

# ON THE HUNGARIAN SUNG VOWELS

**Andrea Deme**

**Department of Phonetics, Eötvös Loránd University,  
Research Institute for Linguistics, Hungarian Academy of Sciences**  
e-mail: deme.andrea@nytud.mta.hu

## **Abstract**

Singing at a very high pitch is associated with vocal tract adjustments in professional western operatic singing. However, as of yet there is an inadequate amount of data available on the extent of the acoustic transformation the Hungarian vowels undergo during singing. The author's purpose is to evaluate the acoustic and articulatory changes of Hungarian vowel qualities, and examine the effect of these changes on the intelligibility of sounds, which has not yet been done for Hungarian. The paper contains a brief summary of formerly described tendencies for other languages and data for Hungarian from pilot studies carried out by the author with an adult soprano's and a child's sung vowels.

## **1 Theoretical introduction and questions**

High-pitched singing in the western operatic style demands special articulatory movements and therefore is a specific object of analysis. Possible vocal tract adjustments one uses while singing have already been extensively described, but the effect of these modifications on the acoustic domain and the perception of sounds can differ considerably from language to language depending on the vowel system. Moreover, characteristics and registers of the high vocal range are less studied, because of its dependence on different techniques and training methods. Thus the aim of the research reported here is to investigate the effect of the articulation of singing on the Hungarian vowels /ɔ, a, ε, e, i, o, ø, u, y:/ on production and perception as well. Assuming that singing and speech can be understood with equal ease, the operatic tradition does not have the practice of subtitling the performances played in the language of the audience (Watson, 2009). Consequently, our research by proving increased difficulty in perceiving the high-pitched sung language elements, might point out the necessary change of this practice.

There is some agreement, that in speech, the vowel can be characterised by its first two formants ( $F_1$ ,  $F_2$ ) (Peterson and Barney, 1952), and these can also be a cue to their perception (Gósy, 1987; Neary, 1989). These resonances in adult speech normally lie far above the speaker's fundamental frequency ( $f_0$ ), but in high-pitched singing, the  $f_0$  is often raised above the average value of  $F_1$  (or occasionally even the  $F_2$ ). In this case, maintaining the normal vowel-dependent values of  $F_1$  would not only change the timbre required for the western operatic style, but the singer should use greater vocal effort (along with unhealthy phonation) to provide the necessary loudness as well. To avoid producing weak sound (i.e. losing timbre and loudness),

when  $f_0$  exceeds  $F_1$  trained singers start to tune  $F_1$  (i.e. adjust the first resonance of the vocal tract) to the value of the raised fundamental (Sundberg, 1989; Garnier et al., 2010), therefore enhancing the amplitude of  $f_0$ . At high pitch this tuning can be described linearly (where  $F_1:f_0$  tuning is controlled by jaw movement indirectly), but in lower ranges it can only be described by more complex nonlinear coupling effects (i.e. impedance-matching) (see Titze and Worley, 2009).

In general at high pitch,  $F_1$  can be tuned by lowering the jaw and unrounding the lips (Sundberg, 1969). The position of the larynx is also changed with ascending pitch, but the direction and the nature of its movement seem to have great variability across singers. According to Sundberg (1969), the vertical position of the larynx in singing is (broadly speaking) inversely proportional to pitch. Hurme and Sonninen (1995) later described four basic movement strategies that can be observed in female and male singers: the larynx can be pulled in an 1) anterior-superior (up and forward), 2) posterior-superior (up and backward), or an 3) inferior (down) direction, but it also can have a 4) complex, zig-zagging route while raising  $f_0$ . Hurme and Sonninen also showed that the cartilages and the hyoid bone can change their “textbook” position to extreme constellations (i.e. the hyoid bone can move in a quite anterior-inferior position to the front of the thyroid cartilage).

The specific articulatory features resulting in changes of the vowel’s formant values, and the raised  $f_0$  associated with wider harmonic spacing (which means limited resolution on conveying the transfer function of the vocal tract) have the effect of reducing the acoustic vowel space with ascending pitch, and producing acoustically similar vowel qualities at the higher vocal range (Scotto di Carlo and Germain, 1985; Dowd et al., 1998; Joliveau et al., 2004; Millhouse and Clermont, 2007; Wolfe et al., 2009). In addition, some research also implied that there seem to be learned relationships between  $f_0$  and the formant frequencies, which support human speech processing to distinguish vowel qualities, so the changes in these relations presumably distract the perceptual mechanisms in some extent (Assmann et al., 2002). At the higher boundary of this tuning (reaching a certain  $f_0$  at about 800–900 Hz, although according to Watson’s description (2009), it already happens at 698 Hz) singers tend to use a single canonical (wide open) vocal tract shape while producing all the vowels (Millhouse and Clermont, 2007; Bresch and Narayanan, 2010), therefore decreasing the distinction of acoustic vowel space not only in perception, but in production as well.

Three earlier studies regarding the perception of the high-pitched sung vowels described the increasing number of errors in identification with raising the fundamental frequency of the singing voice. Gottfried and Chew (1986) revealed that back vowels are more often misidentified than front vowels, and the different phonatory modes (so called “registers”) have an important effect on the perception of the vowel sounds as well. According to Scotto di Carlo and Germain (1985), the vowels not properly identified are mostly rounded and closed, and are generally confused with open and central ones, in particular [a]. Hollien et al. (2000) described these tendencies of confusion as shifts towards vowels with higher  $F_1$  (which

practically means a more opened configuration on the articulatory domain). Besides the latter two papers, no data on tendencies of errors occurring in misidentification were presented. In any other investigations, there are just assumptions posed (based on purely acoustic and articulatory data), implying that due to more open articulation at high pitch, vowels appear as more open in perception, too. However, it is well-known from the literature that production and perception have a non-trivial relationship, and it is a matter of agreement that any supposition of this kind has to be verified perceptually.

Nevertheless, the acoustic and perceptual tendencies of the changes of vowel production presented above are highly language dependent, since the vowel inventory differs among languages. Thus, the purpose of this study is to examine the acoustic and articulatory features of the production of the singing voice for the first time in Hungarian. The main questions are the following: 1) Is it possible to distinguish different vowel qualities in Hungarian at a relatively high pitch? 2) Which are the critical values of  $f_0$  in the perception of sounds produced in the higher ranges of the singing voice? 3) What are the tendencies for identification errors for high-pitched sung vowels and what articulatory background can be hypothesized for these confusions? 4) Which are the critical values of  $f_0$  in the formant tuning with increasing the fundamental? 5) What happens to the vowels' other formants during singing?

The acoustics of children's sung vowels are not well known yet for any language, but for Hungarian, even the acoustics of children's speaking voice is under-researched (see e.g. Gósy, 1984; Deme, 2012b). However, the short vocal folds and the generally smaller vocal tract (therefore higher values of resonances and formants of speaking voice) of children imply the supposition that sung vowels and acoustic vowel space behave differently (from adults) in their singing production (e.g. because of high  $F_1$ , no tuning is necessary in case of /a:/). Moreover, it is also not clear what differences the lack of many years of training can create with regard to the energy of the spectral components of the child's voice (i.e. the vocal efficiency or loudness). To examine these questions, a pilot study was carried out on an 8-year-old girl's sung vowels. In the following sections the results of these studies and further questions are presented.

## **2 Pilot studies**

In the ongoing work, three pilot studies have been carried out this far. In Experiment 1, the formant frequency changes and their effect on speech processing were analysed. Therefore acoustic analysis and perception tests were performed on 9 Hungarian vowels (which can be uttered with an extended duration without changing the vowel quality itself). Since the results were not entirely in agreement with previous findings described for other languages, the effect of consonantal context was tested in Experiment 2. Experiment 3 was a pilot study regarding formant tuning and vowel space reduction in a trained child's singing productions.

## 2.1 Subjects, material and method

As it was demonstrated earlier, vowel identification is better in consonantal context (Strange and Verbrugge, 1976). On the other hand (with regard to singing), high pitch can be achieved the most easily (without “forcing”) while pronouncing vowels in vowel-like (e.g. nasal) context (Kerényi, 1959). Therefore for Experiment 1, nonsense *mVn* utterances were recorded, where the vowels /ɔ, a, ε, e, i, o, ø, u, y/ in a nasal context were sung by a professional soprano singer (age 50) at a comfortable loudness, at the steady-state  $f_0$  values of 500, 550 and 650 Hz. For the acoustic analysis, the singer’s spoken vowels ( $f_0 \sim 200$  Hz) were used as a reference. During the perception test, the subjects’ (4 males, 6 females) task was to listen to the presented sequences, and fill in the blanks left for the vowel between the given consonants on an answer sheet. Vowels in other consonantal contexts were also recorded and presented as distractor stimuli. For analysis, 36 sequences were used (1 context  $\times$  9 vowels  $\times$  4 fundamental frequencies =).

Since in the first study, disagreements were found with the earlier demonstrated confusion tendencies, the second experiment did not include just nasals, but voiced and unvoiced fricatives /ʃ, ʒ/ were also recorded (in *mVn*, *sVs*, *zsVzs* sequences) at the fundamental values of 500, 550, 600, 650 Hz and in speech produced by the singer from Experiment 1. The listeners (10 females, 5 males) (after hearing the whole stimulus) were asked to click on the vowel’s orthographical symbol displayed on the computer screen. For analysis, 270 sequences were used (3 context  $\times$  9 vowels  $\times$  5 fundamental frequencies  $\times$  2 repetitions =). In both cases, the listeners were non-trained subjects, since we wanted to demonstrate the case of an average member of the opera audience.

In the third study, an 8-year-old girl’s sung and spoken /a, i, u/ vowels (in *IV* context) and three folk songs were recorded. The participating child was attending music school, and she was at the beginning of her training. The *IV* sequences were uttered in an ascending and descending scale from F3 (175 Hz) to F5 (698 Hz)<sup>1</sup>. For the acoustic analysis, we recorded a 30-year-old soprano’s vowels for comparison. It is an important thing to emphasize, that this time, equally tempered musical notes were used (and not just a scale with physically equidistant frequencies with no regard to musical conventions or hearing).

All of the recordings were carried out in a soundtreated room, and digitized at 44.1 kHz. The formants were determined by Fourier analysis (FFT) using Praat (Boersma and Weenink, 2011) and Wavesurfer (Sjölander and Beskow, 2009) at the middle of the vowel duration. Considering the difficulties of estimating formant frequencies from the output signal at high pitch (see 1<sup>st</sup> sec. 4<sup>th</sup> par.), it has to be

<sup>1</sup> In this paper musical notes are referred to by their conventional musical names according to the Acoustical Society of America (Young, 1939). To clearly distinguish between formant frequencies and pitch values, we refer to formants by numbers in index (e.g. F<sub>2</sub>) and to musical notes with numbers of normal size (e.g. F2).

emphasized that the presented formant values measured in these cases can only refer to the frequency of the enhanced harmonics (that possibly lie in the bandwidth of the corresponding formant), and may not be the center of the formant in question. (Similar restrictions are required when interpreting the results of Hollien et al., 2000) The listening tests were presented under headphones.

## 2.2 Results

### 2.2.1 Experiment 1

In Experiment 1 (Deme, 2012a), the results of the perception test showed a non-monotonic yet descending trend for correct identification percentages with ascending  $f_0$  (Fig. 1). This means there is a reduction of intelligibility proportional to pitch. The cause of the sudden peak at 550 Hz is not clear yet, but it might be the effect of the so-called “register transition”. The registers divide the tonal scale into pitch intervals, which are produced with the same phonatory mode, but at register transitions, changes in phonation can be observed (Titze, 2008). It was revealed that while the identification rate decays at the highest portion of a register, reaching the next (upper) one (with more optimal phonatory position) causes improvement in maintaining vowel intelligibility (Scotto di Carlo and Germain, 1985). Thus, a drop and jump in the identification rates can be the mark of a switch in production mode while raising pitch. Since there is no acceptable agreement on its acoustic properties, the effect of transition cannot be analyzed directly from objective acoustic data. One way of assessing its presence can be perceptual assessment carried out with trained listeners (singers, singing teachers), which is planned to be done in the future.

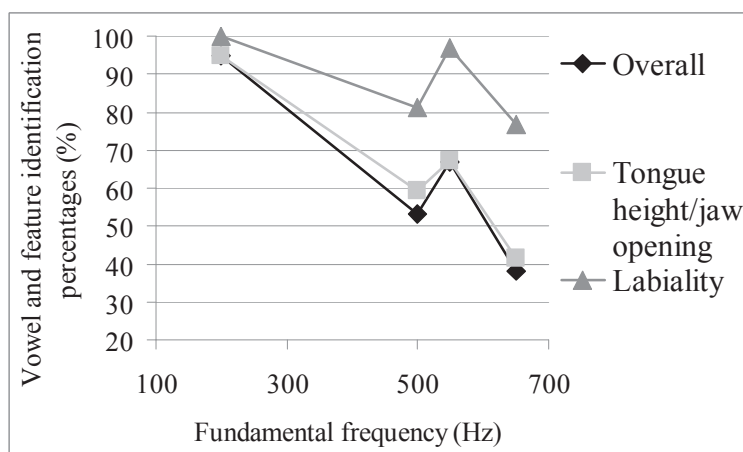


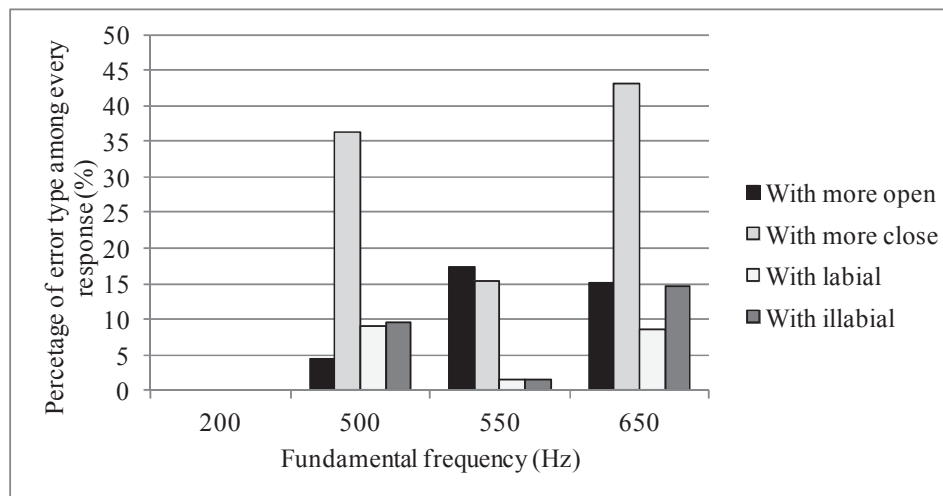
Figure 1. Vowel and feature<sup>2</sup> identification percentages in the function of fundamental frequency.

<sup>2</sup> Examples: along the [close] dimension /u:/ is close, /o:/ is close-mid, /ɔ:/ is open-mid, /a:/ is open; along the [labial] dimension /o:/ and /u:/ are labial, /e:/ and /i:/ are non-labial.

As it is seen in Fig.1 labiality is more resistant to pitch than tongue height ( $F(2) = 8.34$ ;  $p = 0.02$ ), which is inconsistent with data for French (Scotto di Carlo and Germain, 1985). It also can be seen that the percentages of correct perception of the feature [close] are much higher when the vowels are spoken or produced at the lower  $f_0$  values of singing (500 and 550 Hz). This finding is not surprising, as numerous studies have already shown that jaw opening is inversely proportional to  $f_0$  (e.g. Sundberg, 1969, 1987; Austin, 2005; Bresch and Narayanan, 2010). That is, the higher the  $f_0$  is, the lower the tongue/jaw is positioned. However, the types of confusions, the vowel qualities involved, and the percentages of correct recognition in detail practically show the opposite of the two already available descriptions as demonstrated in Table 1 and Fig 2.

*Table 1.* Percentages of correct identification per vowel per fundamental frequency.

$f_0$	Recognition percentages								
	a:	ɔ	o:	u:	ø:	y:	ε	e:	i:
Speech (~200 Hz)	100%	90%	100%	95%	100%	100%	100%	100%	71%
500 Hz	22%	77%	62%	85%	4%	50%	73%	24%	90%
550 Hz	100%	71%	55%	43%	50%	71%	95%	60%	58%
650 Hz	26%	9%	30%	43%	0%	65%	67%	38%	67%



*Figure 2.* Confusions in vowel identification per fundamental frequency. Every response for each stimulus was worth two judgements (for the two features: [labial], [close]).

Contrary to what was demonstrated by Scotto di Carlo and Germain (1985) and Hollien et al. (2000), here at high pitch, the highest recognition percentages were measured at close (/y:, i:/) and open-mid (/ε/) vowels, and the vowel with the widest jaw opening (/a:/) could only maintain its intelligibility for a smaller extent (with a high recognition rate at 550 Hz). At the highest fundamental, the recognition

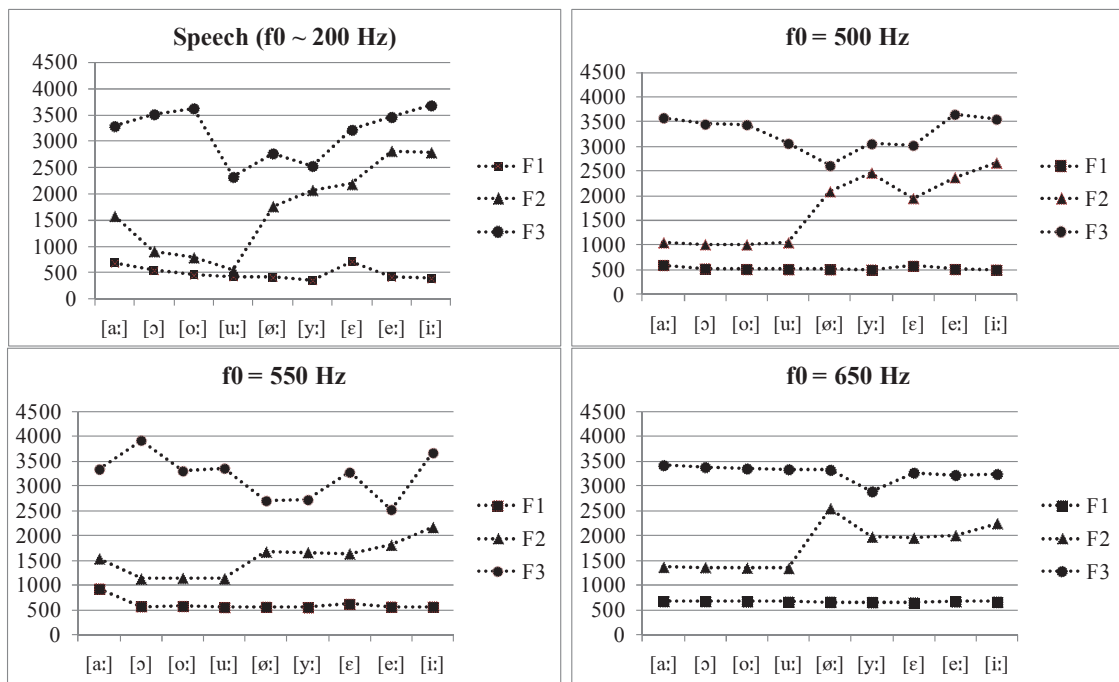


percentages of the feature [close] were 58.5% for close, 22.7% for close-mid, 37.2% for open-mid vowels and 26.1% for the open /a/.

To analyze the subjects' responses, confusion matrices were constructed: one matrix per each fundamental frequency (for example see Table 2.). As the matrices show not only the number of confusions, but the types of errors that occur as well, this can be an efficient way of summing up the results.

*Table 2.* Example of a confusion matrix at the highest sung fundamental frequency ( $f_0 = 650$  Hz). Number of occurrences of the stimulus/response pair is indicated in the corresponding box of the matrix. For example: the vowel [a:] was mistaken for [ɔ] 12 times, for [o:] twice, and for [u:] 3 times at 650 Hz.

		Response								
		a:	ɔ	o:	u:	ø:	y:	ɛ	e:	i:
Stimulus	a:	6	12	2	3					
	ɔ		2	7	13					
	o:		3	6	11					
	u:		9	3	9					
	ø:					0	1	2	4	15
	y:						15		5	3
	ɛ							14	3	4
	e:							1	9	14
	i:								7	14



*Figure 3.* Measured first three formant frequencies of each vowel per fundamental frequency.

Despite the assumption, that the singer tends to articulate close and labial vowels as more open and non-labial with increasing pitch, the greatest proportion of error types in perception seems to be somewhat the opposite: not properly perceived vowels tended to be indentified as more closed sounds, in particular as /i:/ (27% out of the total number of mistaken vowels) (Fig. 2). The second most frequent vowel in the hierarchy of mistakes was /ɔ/ (16%), which might be a language or speaker specific articulatory feature of singing but still meets the expectation for vowel production with opened jaw. But as the most close vowel, /i:/ cannot fit the notion of earlier demonstrated articulation tendencies in any way.

In agreement with previous studies, the acoustic (i.e. frequency structure) differences of distinct vowel qualities are reduced with ascending pitch (Fig. 3). As the  $f_0$  reaches the average value of  $F_1$ , the  $F_1$  can not be distinguished from the raised pitch any more; this becomes common through the whole vowel spectrum<sup>3</sup>. (For certain cases (e.g. in the case of /ɛ/) the tuning seems to begin before the  $f_0$  would exceed the vowel's  $F_1$ .) Since, the first spectral maxima and the  $f_0$  coincide on the vowel's spectrum at any sung pitch, it can be assumed, that the singer tends to shift  $F_1$  to match  $f_0$  as expected. This tuning implies increase of the jaw opening with  $f_0$  on the articulatory domain. However, the effect of more open articulation was not found in the results of the perception test.

It seems that at the higher sung  $f_0$ s,  $F_2$  still remains as a cue for vowel frontness, that is, the back–front distinction seems to be the most resistant of all the articulatory and acoustic changes the vowels undergo. Not finding any example of back–front confusions in identifying the vowels at the examined  $f_0$  scale confirms this observation. (Note that according to the quantal theory, the  $Sg_2$  occurring at about 1400-1600 Hz for adult speakers is a natural separator for this feature. [see Stevens 1989 and Section 3 of the recent paper.]

As hypothesized, reduction and a categorical shift towards the vowel quality of [a:] can be observed on the acoustic vowel space (consisting of the most spaced vowels /a:, i:, u:/ in  $F_1 \times F_2$  domain) (Fig. 4).

Considering the remarkable decay in acoustic vowel-differentiation with ascending pitch, the low recognition percentage of overall vowel identification (38%) at the highest pitch is not surprising. However, the appearance of the most dominant type of error (mistakes for /i:/, 27%) are not sufficiently supported by the acoustics. (As for the back vowels the high percentage of confusions with /ɔ/ [16%] practically meet the assumptions of mistakes for /a:/.)

Since the acoustic data coincide with earlier description, but in the perception tests some disagreements were found (precisely the types of errors and vowel qualities involved), in Experiment 2 the contextual effect on the intelligibility of

<sup>3</sup> However, considering the known limitations of the FFT analysis, it can not be excluded that the tuning might make  $F_1$  to appear slightly higher than the  $f_0$ , as some of the authors suggest (i.e. Titze and Worley, 2009).



sung vowels was tested (Deme, 2011a). It was also a question, whether the nasal context of a sung vowel decreases the intelligibility as suggested before (Rosner and Pickering, 1994).

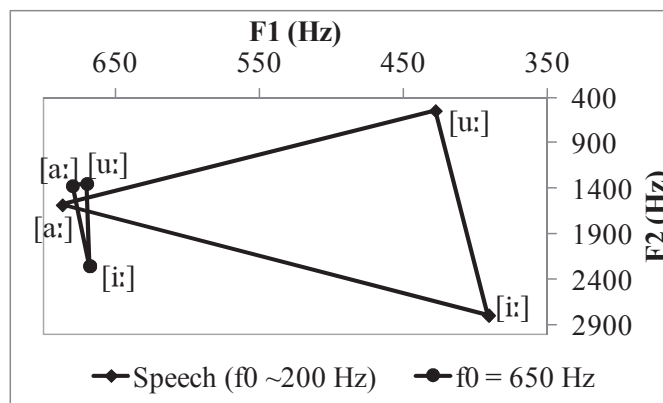


Figure 4. The acoustic vowel space consisting of /a:/, i:/, u:/ in speech and in singing at the highest fundamental frequency.

### 2.2.2 Experiment 2

In Experiment 2, the recognition percentages decreased monotonically (fell beneath 50% at 550 Hz) and appeared to be the highest for /a:/ on each fundamental frequency. The overall rate of correct identification of this vowel was 78%. This finding fits the assumption, that the opening of the jaw is enhanced while singing, therefore the intelligibility of open and mid-open vowels are the easiest to maintain even at higher pitch. At the same time, the most frequent confusion occurring during misidentification of vowels was neither with the presumed /a:/, nor /ɔ:/ or /i:/ as observed in Experiment 1, but with /y:/ (24% of all mistakes) followed by /ø:/ (18%) (in particular in the case of /ø:/, i:/, e:/, ε/).

In voiced and unvoiced fricative contexts practically the same tendency appeared as in nasal contexts: the singer mostly tended to articulate vowels which were perceived as having a more closed jaw instead of increasing the opening (Fig. 5, 6). Labiality-related confusions were less frequent than those concerning jaw opening.

Confusions with more closed sounds were dominant in every context, and at every fundamental frequency investigated. Thus, no differences were found in the function of pitch or context in this domain.

To inspect the contextual effect on vowel intelligibility, we totaled the number of confusions per consonantal environment (Fig 6), and ran the  $\chi^2$  test to evaluate the degree of differences. The statistical analysis showed significant deflection among the three types of carrier sequences ( $\chi^2 = 8.511$ ,  $df = 2$ ,  $p = 0.014$ ), but as it is seen in Fig. 6, the highest identification rate characterized not the fricative, but the nasal context. The result can be interpreted as it is not only easier to sing vowels between nasal consonants at a relatively high pitch, but also efficient with regard to the matter of maintaining distinct vowel quality.

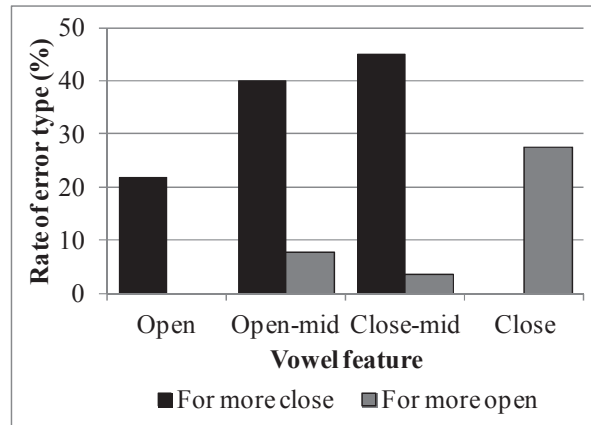


Figure 5. Percentages of error types concerning jaw opening in the function of the vowel feature [close].

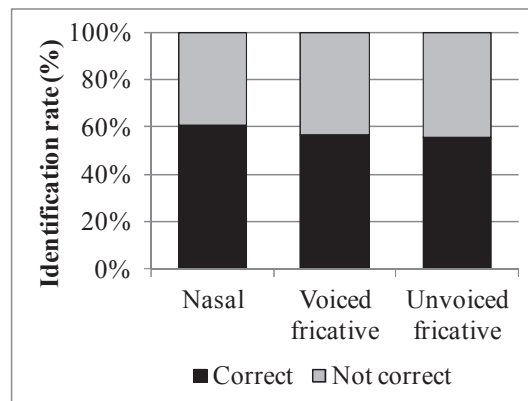


Figure 6. Correct vowel identification in terms of consonantal environment types

### 2.2.1 Experiment 3

In Experiment 3 (Deme, 2011b), an 8-year-old girl's sung vowels were studied. The vowel space measured in speech was greater (together with higher  $F_1$  and  $F_2$ ) than in the case of the adult woman (Fig. 7).

The sung material revealed that no vowel space reduction occurred at higher fundamentals in the child's singing (as expected from adult's data), and even a slight increase in vowel spacing was present. This change may be the result of more open articulation in the case of /a:/ and more fronting in the case of /a:, i:/. Since it was not possible any more to distinguish  $F_1$  from  $f_0$  for /i:/ and /u:/ at A4 (440 Hz), it can be assumed that  $F_1:f_0$  tuning seems to begin at this fundamental frequency, but no direct tuning appeared while producing /a:/, since its average  $F_1$  lies far above the  $f_0$  values used by the child in singing. Therefore, the distinction among vowels seems to be preserved in the acoustic domain at higher registers as well, but (so far) there is no evidence for this separation from a perceptual point of view.

Preliminary results suggested that children's sung vowels contain less energy in the higher frequencies, so the power of the voice might be more limited in loudness than in adults' singing activity. Since this would obviously have serious

consequences on training and learning methods, it needs proper validation in the future.

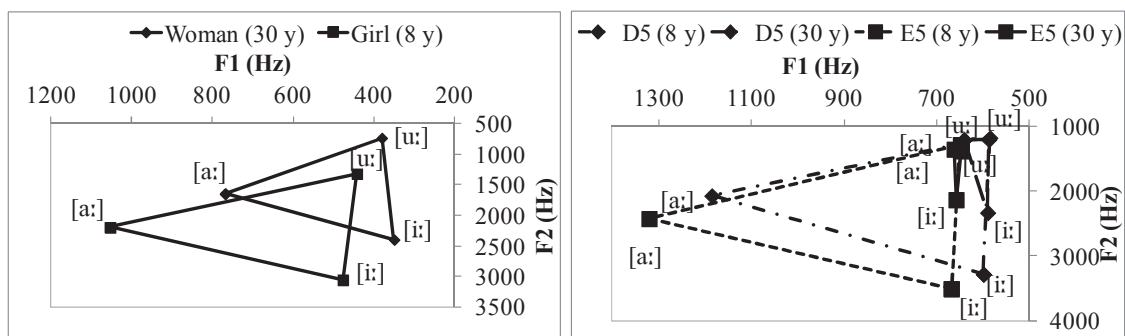


Figure 7. The acoustic vowel space of an 8 year old girl compared to a 30 year old soprano in speech (left), and at D5 (= 588 Hz) and E5 (= 659 Hz) sung fundamentals (right)

### 3 Discussion and future work

The aim of the author's ongoing PhD project is to determine the effect of the articulation of singing on Hungarian vowels in both the acoustic and the perceptual domains (and – to a certain extent – the effect of age). The three experiments carried out so far answered the previously formulated questions in the following way.

1) In Experiment 1 the reduction of vowel space and assimilation of vowel qualities was found to be in agreement with previous studies for other languages. However, total reduction of perceptual vowel differentiation was not observable in our material.

2) The rate of correct recognition decreased gradually with ascending pitch. The recognition percentage for the vowels produced at a relatively high pitch (650 Hz) was 38% for Experiment 1 and 37% for Experiment 2.

3) There was a tendency for the misidentification of vowels with more closed sounds (in particular as /i:/ in Experiment 1 and /y:/ in Experiment 2). This result is partly inconsistent with earlier implied tendencies, and could suggest a more closed articulation in the higher registers (in particular, in the case of front vowels). In Experiment 1 at the highest pitch, it was not open, but close sounds (/y:, i:/) that maintained their intelligibility the most. In Experiment 2, the influence of nasal context on the error types occurring in sung vowels' perception was hypothesized, but it was not confirmed. (Moreover, even some positive effects of nasal environments on vowel intelligibility were demonstrated. This finding supports the empirically established habit of singers for often using nasals for vocal exercising and training.) In this study, the error types and sustainability of vowel features seemed to be much more in line with previous research. Though the seemingly unexpected results need further parsing, possible explanations can be formulated already. Some of the inconsistencies found can be the consequence of the Hungarian vowel system. As an example, it is easy to see that /ɔ/ having three minimal pairs

differing in just one articulatory feature can dominate the hierarchy of errors much more than the expected /a:/ (in Experiment 2), which has none (as it differs from other vowels in two features or even more stages of “closedness”). In a language that has more front vowels than back ones, it is also obvious that this ratio will be represented by the higher number of errors for front vowels as well (at least at relatively lower pitches where there is no total vowel space reduction yet). The high number of mistakes for more closed sounds might be accounted for by the influence of child voice. Since the  $f_0$  of children’s speech and the corresponding formant values are normally higher than in adults’ vowels, it is possible that the practice an average listener has in perceiving high-pitched sounds makes the processing system expect the formant values to be high as well. However, despite raising some of the formants observable in singing, adults’ sung vowels can never have as high frequencies as it would be expected in child speech. Thus the high  $f_0$  accompanied by relatively lower formants can cause the impression of more closed and more back articulation – while the objective acoustic data show no sign for these tendencies. This is possibly the case in the high number of mistakes for the closed vowel /y:/ in Experiment 2 or in the high recognition percentages for /y:/ and /i:/ in Experiment 1 and for /u:, y:/ in Experiment 2. Last but not least, the limitations of the results quoted from the reference literature should be noted. First, the study by Scotto di Carlo and Germain (1985) used only one singer and the same amount of material as the present one, which obviously means that the inconsistencies of their results and those presented here lie not only in the setting of this investigation, but might likely be the result of their design. Second, the paper written by Hollien et al. (2000) did not publish any exact data on the tendencies in question, just reported on some agreement with previous findings of Scotto di Carlo and Germain. Since no percentages or numbers are given in detail, no exact comparison of results is possible, and the report can only be handled with reservations. As for the childrens’ sung vowels, according to our results, no vowel space reduction occurred in the acoustic domain. Although the results suggest greater vowel distinction for high-pitched singing, it still needs perceptual validation along with the issue of vocal power, which seems to be limited for children and thus can have serious consequences for the methods of training.

4) The formant tuning in adult singing began from the lowest sung fundamental frequency studied (500 Hz), and caused acoustically similar qualities within the groups of back and front vowels, but complete vowel space reduction was not reached by  $f_0 = 650$  Hz, therefore no back–front mistakes were found. As for the child, no direct  $F_1:f_0$  tuning was found for /a:/, but it was observable in other vowels (with lower  $F_1$ ) at the pitch value of G4 (392 Hz) and above. Along with less vocal efficiency, the features of tuning (critical  $f_0$  and tendencies) seem to be two of the most important differences between the adult and child performers.

5) As for the higher formants, the predicted formant compression and assimilation within front and back vowel groups was observable.

In addition to the above-mentioned topics, several further questions arose during the investigation. One of these is the problem posed by the appearance of the subglottal resonances in the production of sung vowels. According to Stevens' quantal theory (1989), the subglottal system, having its own resonances ( $Sg_1$ ,  $Sg_2$ ...) influences the speech signal, which has (among others) an important role in supporting the distinctive features of speech sounds. As an example,  $Sg_2$  is believed to be a cause of natural separation between back and front vowels, as  $Sg_1$  is between high and low vowels. Since normally we have no direct muscle control over our subglottal system, the values of  $Sg$  frequencies are roughly constant. However, as we already know, in singing the height and shape of the larynx (and the tension of the vocal folds) can change considerably, therefore the values and role of  $Sg$  seems to be problematic. Do  $Sg_1$  and  $Sg_2$  maintain their natural frequencies or function in vowel differentiation at higher  $f_0$  as well? As it was mentioned in Experiment 1, in view of the formant structure of back and front vowel groups, we suppose so. Our preliminary results (Grácsi and Deme, 2011) show that although measuring the  $Sg$  values is almost as difficult as measuring the formants at higher registers, it seems that  $Sg_1$  and  $Sg_2$  slightly shift upwards with ascending  $f_0$  (probably as a result of adjusting the larynx height or changing the phonatory position of the vocal folds), therefore its distinctive function supposedly can remain (to a certain degree) in the higher registers as well.

As the studies show, vocal tract adjustments observed in singing often make vowels ambiguous, but relatively invariant information provided by consonant articulation still can make the sung text intelligible. Therefore accurate consonant articulation in training of singing seems to be a significant factor to enhance.

## References

- Austin, S.F. 2005. Jaw opening in novice and experienced classically trained singers. *Journal of Voice* 21(1), 72-79.
- Assmann, P. F., Nearey, T. M. and Scott, J. M. 2002. Modeling the perception of frequency-shifted vowels. *Proceedings of the 7<sup>th</sup> International Conference of Spoken Language Processing*. 425-428.
- Boersma, P. and Weenink, D. 2012. Praat [Computer program] (Version 5.3). <http://www.praat.org>, accessed on 25 May 2012.
- Bresch, E. and Narayanan, E. 2010. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. *Journal of the Acoustical Society of America* 128(5), 335-341.
- Deme, A. 2011a. Az énekelt magánhangzók észlelése réshangkörnyezetben. [Perception of sung vowels uttered in fricative context.] In Váradi, T. (szerk.) *V. Alkalmazott Nyelvészeti Doktoranduszkonferencia*, Budapest: MTA Nyelvtudományi Intézet. 16-28.
- Deme, A. 2011b. Egy nyolcéves gyermek énekelt és beszélt magánhangzóinak akusztikai jellemzői. Esettanulmány. [Acoustic features of an 8-year-old girl's sung vowels.] *Alkalmazott Nyelvtudomány* 11(1-2), 169-188.
- Deme, A. 2012a. Az énekelt magánhangzók fonetikai elemzése. [Phonetic analysis of hungarian sung vowels.] In Parapatics, A. (szerk.) *Doktoranduszok a nyelvtudomány útjain. A 6. félévű konferencia*, Budapest: ELTE Eötvös Kiadó. 33-46.
- Deme, A. 2012b. Óvodások magánhangzóinak akusztikai jellemzői. [An Acoustic Analysis of Vowels in 6-7-year-old Hungarian Children's Speech] In Markó, A. (szerk.)



- Beszédtudomány: Az anyanyelv-elsajátítástól a zöngékezdesi időig.* [Speech Science: From Language Acquisition to Voice Onset Time.] Budapest: ELTE és MTA Nyelvtudományi Intézete. 77-99.
- Dowd, A., Smith, J.R. and Wolfe, J. 1998. Learning to pronounce vowel sounds in a foreign language using acoustic measurements of the vocal tract as feedback in real time. *Language and Speech* 41, 1-20.
- Garnier, M., Henrich N., Smith J. and Wolfe J. 2010. Vocal tract adjustments in the high soprano range. *Journal of the Acoustical Society of America*. Vol. 127. No. 6, 3771-3780.
- Gósy, M. 1984. Hangtani és szótani vizsgálatok egy hároméves gyermek nyelvén. [Studies on phonetical and lexicological features in 3-year-old children's speech.] *Nyelvtudományi Értekezések* 102. Budapest: Akadémiai Kiadó. 19-42.
- Gósy, M. 1987. A formánsszerkezet változásának hatása a magánhangzók felismerésére. [The effect of changes in the formant structure of vowels on vowel perception.] *Magyar Nyelv* VOL LXXXII No. 2. 49-59.
- Gottfried, T. L. and Chew, S. L. 1986. Intelligibility of vowels sung by a countertenor. *Journal of the Acoustical Society of America* 79(1), 124-130.
- Gráci, T. E. and Deme, A. 2011. A szubglottális rezonanciák megjelenése az éneklésben. (talk) [Subglottal resonances in singing.] *XIII. Pszicholingvisztikai Nyári Egyetem*. Balatonalmádi, 22–26 May 2011.
- Hollien, H., Mendes-Schwartz, A. P., and Nielsen, K. 2000. Perceptual confusions of high-pitched sung vowels. *Journal of Voice* 14(2), 287-298.
- Hurme, P. and A. Sonninen. 1995. Vertical and sagittal position of the larynx in singing. In Elenius, K. and Branderud, P. (eds) *Proceedings of the XIII International Congress of Phonetic Sciences*. 214-217.
- Joliveau, E., Smith, J. and Wolf, J. 2004. Vocal tract resonances in singing: the soprano voice. *Journal of the Acoustical Society of America* 116(4), 2434-2439.
- Kerényi, M. Gy. 1959. *Az éneklés művészete és pedagógiája*. [The Art and Teaching of Singing.] Budapest: Zeneműkiadó.
- Millhouse, T. and Clermont, F. 2007. Acoustic description of a soprano's vowels based on perceptual linear prediction. *16th International Congress of Phonetic Sciences*, Saarbrücken, 6-10 August, 2007. 901-904.
- Near, T. M. 1989. Static, dynamic, and relational properties in vowel perception, *Journal of the Acoustical Society of America* 85, 2088-2113.
- Peterson, G. E. and Barney, H. L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24(2), 175-184.
- Rosner, B. S. and Pickering, J. B. 1994. *Vowel Perception and Production*. Oxford: Oxford University Press.
- Scotto di Carlo, N. and Germain, A. 1985. A perceptual study of the influence of pitch on the intelligibility of sung vowels. *Phonetica*, 42(4), 188-197.
- Sjölander, K. and Beskow, J. 2009. Wavesurfer [Computer program] (Version 1.8.8). <http://www.speech.kth.se/wavesurfer>, accessed on 25 May 2012.
- Stevens, K. N. 1989. On the quantal nature of speech. *Journal of Phonetics* Vol. 17. 3-46.
- Strange, W. and Verbrugge, R. R. 1976. Consonant environment specifies vowel identity. *Journal of the Acoustical Society of America* 60(1), 213-224.
- Sundberg, J. 1969. Articulatory differences between spoken and sung vowels in singers. *STL-QPSR* 10(1), 33-46.
- Sundberg, J. 1987. *The Science of the Singing Voice*. Illinois: Northern Illinois University Press.
- Titze, I. R. and Worley, A. S. 2009. Modeling source-filter interaction in belting and high-pitched operatic male singing. *Journal of the Acoustical Society of America*. 126(3), 1530-1540.
- Titze, I. R. 2008: Nonlinear source-filter interaction in phonation – theory. *Journal of the Acoustical Society of America* 123(5), 2733-2749.



- Watson, A. H. D. 2009. *The Biology of Musical Performance and Performance Related Injury*. Lanham: Scarecrow Press.
- Wolfe, J., Garnier, M., and Smith, J. 2009. Vocal tract resonances in speech, singing and playing musical instruments. *Human Frontier Science Program Journal* 3, 6-23.
- Young, R. W. 1939. Terminology for logarithmic frequency units. *The Journal of the Acoustical Society of America* 11(1), 134-139.